Interpreting, Improving, and Augmenting Multi-Model Ensembles

James A. Hansen
MIT, EAPS, 54-1616, 77 Massachusetts Ave, Cambridge, MA 02139
phone: (617) 452-3382 fax: (617) 253-8298 email: jhansen@mit.edu

Grant Number: N000140210473
http://wind.mit.edu/~hansen/

LONG-TERM GOALS

Develop methods to intelligently add new ensemble members to multi-model ensemble forecasts, to maximally exploit existing multi-model ensemble forecasts, and to diagnose model inadequacies and differences through multi-model ensemble forecasts.

OBJECTIVES

This project has two primary objectives.

1. Exploiting existing multi-model ensemble analyses and forecasts

Extract as much information as possible from the analyses and forecasts currently available from different operational Numerical Weather Prediction (NWP) centers.

2. Use of a single model structure to augment and interpret multi-model results

Adjust the parameters of a single, simplified model to mimic the output of the more complex NWP models, and exploit the resulting parametric information for ensemble augmentation and for interpretation of the differences between the models making up the ensemble.

APPROACH

1. Exploiting existing multi-model ensemble analyses and forecasts

It is operational impossible to maintain a multi-model development, data assimilation, and forecasting system at a single NWP center. This motivates extracting as much information as possible from the analyses and forecasts currently available from different operational NWP centers. This collection of analyses and forecasts from different NWP centers is denoted the poor man's multi-model (PM MM) ensemble. Because the existing PM MM has few members (there are only a handful of operational NWP centers around the world), methods for extracting as much information as possible from the ensemble are of interest. To increase the effective ensemble size without adding additional models, this project will explore implementing a lagged average forecasting technique where forecasts launched at different times are combined at common verification times. Because forecasts at longer leads lack the observational information available to short lead forecasts, the ensemble transform Kalman filter (ET KF) (Bishop *et al*, 2001) will be utilized to incorporate observations into existing

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **30 SEP 2002**	2. REPORT TYPE	3. DATES COVERED **00-00-2002 to 00-00-2002**

4. TITLE AND SUBTITLE **Interpreting, Improving, and Augmenting Multi-Model Ensembles**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **MIT, EAPS, 54-1616,,77 Massachusetts Ave,,Cambridge,,MA, 02139**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Develop methods to intelligently add new ensemble members to multi-model ensemble forecasts, to maximally exploit existing multi-model ensemble forecasts, and to diagnose model inadequacies and differences through multi-model ensemble forecasts.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**7**	

ensemble forecasts. In this way each forecasts ensemble member will be conditioned on the same amount of information, regardless of its lead time.

2. Use of a single model structure to augment and interpret multi-model results

Designing an atmospheric GCM from scratch with the aim of optimally augmenting an existing PM MM ensemble is far beyond the scope of this project. Instead, a single (simple) model structure will be used to model the output of the more complex PM MM ensemble members. A given set of PM MM ensemble analysis can be used to determine the simple model parameter perturbations necessary to produce simple ensemble forecasts constrained to lie in the subspace spanned by the PM MM ensemble forecasts. The existing PM MM ensemble can then be augmented by perturbing the simple model's parameters in the direction of these "parametric singular vectors" and produce model states that expand the PM MM ensemble distribution. In addition, insight into the difference between the models in the PM MM ensemble will be gained by examining the required parametric perturbations.

WORK COMPLETED

Preliminary experiments have been performed in an effort to understand model error in a multi-model context. A database of multi-model ensemble forecasts has been obtained from NCAR (National Center for Atmospheric Research) that includes forecasts from the then-NMC (now NCEP, National Centers for Environmental Prediction), the ECMWF (European Centre for Medium-Range Weather Forecasts), and the NCAR CCM3 model from the dates December 1, 1995 through February 14, 1996. Ensemble forecasts of varying sizes and for varying leads for each of the models are launched throughout the period, although all NMC and CCM3 forecasts are initiated from 00Z, while the ECMWF forecasts are initiated from 12Z. The ECMWF forecasts are available, but not the ECMWF analyses.

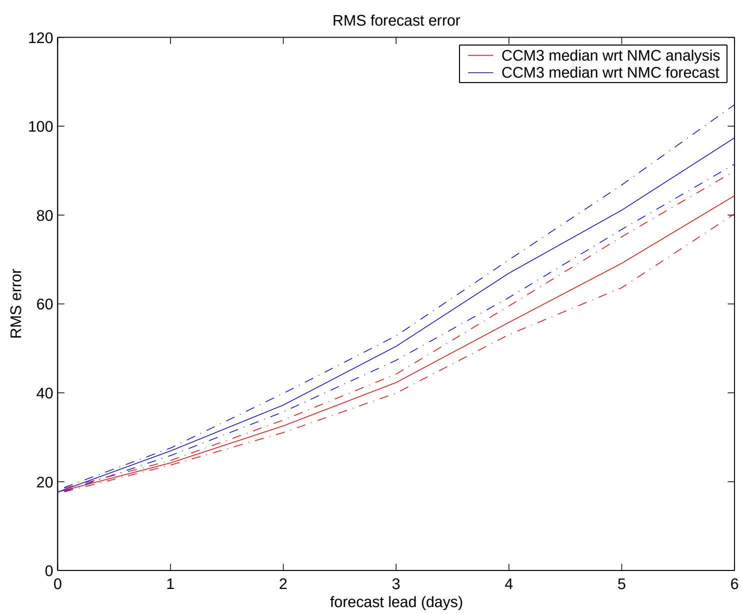

Figure 1: 500mb height median (solid) and one standard deviation (dashed) errors for the CCM3 model versus "truth" as measured by the NMC analyses (blue) and versus the associated NMC forecast (red). It is seen that CCM3 is a better model of the NMC forecast than it is of truth.

It is found that CCM3 is a better model of the NMC forecast than it is of truth (as measured by analyses, see Figure 1). This is consistent with the Hansen (2002) interpretation of the Richardson (2001) multi-analysis ensemble results. Richardson found that much of the benefit of a PM MM ensemble could be replicated by using a single model, but launching ensemble members from the multi-model analyses. One interpretation of this result is that it is not the multiple models that are important, but rather the fact that the different model analyses more effectively sample from initial condition uncertainty space. An alternative interpretation is that the single model is a good model of the other models in the PM MM over short time scales. The fact that CCM3 does a better job mimicking NMC forecasts than mimicking truth supports the latter interpretation, although it certainly not sufficient proof.

Assessing PM MM ensembles proves an interesting problem. The familiar rank histogram approach is utilized, as is the more novel minimum spanning tree (MST) rank histogram. In short, the MST length is the length of the segments that join a collection of points in state space such that the length of the segments is minimized. In the traditional rank histogram, a scalar measure (temperature at a location, say) is taken from each ensemble member and used to form the boundaries of equal probability bins. If the ensemble is drawn from the same distribution as truth, then the scalar verification is equally likely to fall between any two ensemble members. Assessing over a number of different forecasts should lead to a uniform rank histogram if the ensembles are probabilistically correct. A similar approach is taken with the MST rank histograms. The boundaries of equal probability bins are determined by systematically replacing one ensemble member at a time with verification and calculating the associated MST length. The bins are populated by the MST length of the ensemble alone.

When assessing PM MM ensemble forecasts, the MST rank histograms prove far more sensitive than the traditional rank histograms. The PM MM ensemble consists of 11 NMC ensemble members and 10 CCM3 ensemble members. Because the available ECMWF forecasts were launched at 12Z instead of 00Z it was not possible to include them in the PM MM. Taking NMC analyses as verification, the traditional rank histograms for the PM MM ensemble forecasts are statistically indistinguishable from uniform distributions even out to 5 day lead times (although the sample size is very small) (see first row of figure 2). By contrast, the MST rank histograms clearly show that the ensemble is deficient after only a 2 day lead (see second row of figure 2). Assessing the same ensemble forecasts using the CCM3 analyses degrades the results further (third row of figure 2). The CCM3 model does not have its own data assimilation system. Instead, CCM3 forecasts were initiated from the NMC analyses projected into the CCM3 space (and balanced). To further explore the sensitivity of the results to the choice of verification, the single model NMC ensemble forecasts are assessed using first the NMC analyses, and then the CCM3 analyses. The two choices of verification lead to qualitatively different probabilistic assessment results. The NMC and CCM3 analyses only differ by a projection operator, yet utilizing them in ensemble verification implies significantly different ensemble quality. In the single model ensemble case it seems clear that one wants to use that model's analysis as verification, but in the multi-model context it is not clear which verification is "best".

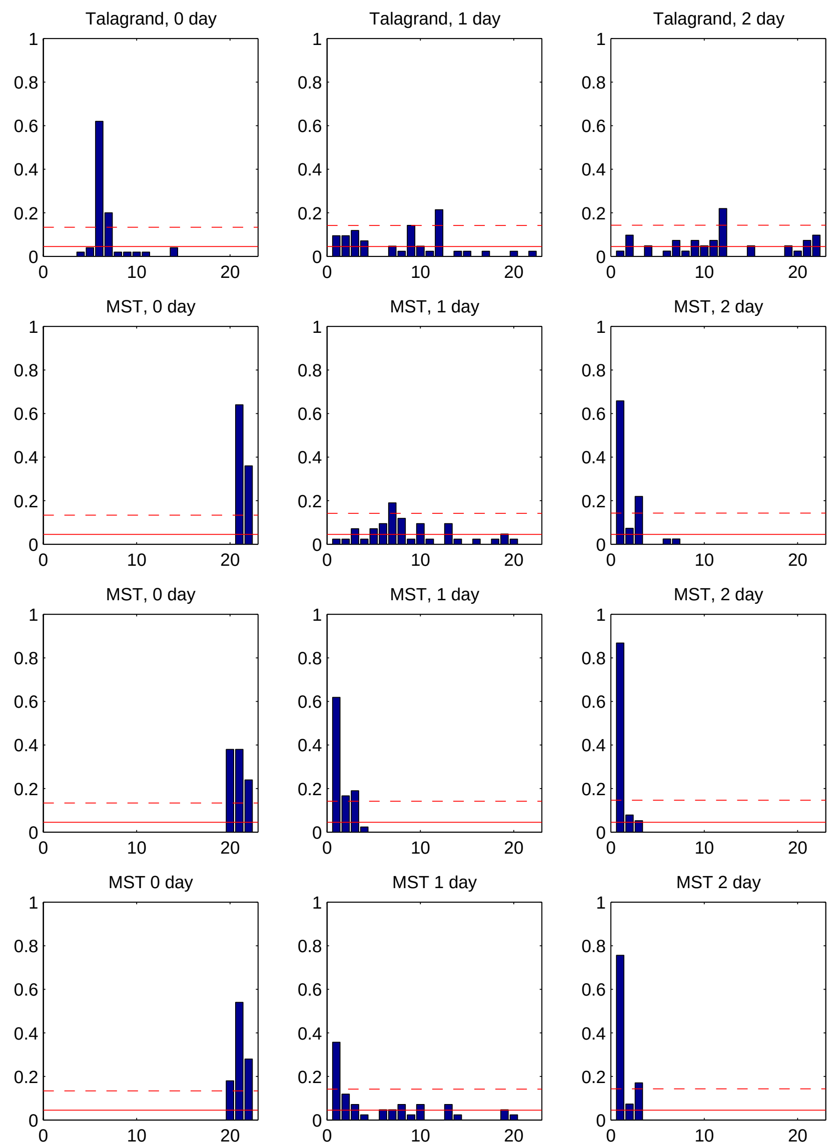

Figure 2: Traditional and MST rank histograms for the PM MM ensemble forecasts. The solid red lines are the expected mean value in each bin, and the dashed red lines are the expected standard deviation. Each column is a different lead (0, 1, and 2 days). The first row is the traditional rank histogram with the NMC analyses as verification, the second row is the associated MST rank histogram. Notice that the MST rank histogram indicates there are problems with the PM MM after only two days, while the traditional rank histogram suggests there are no problems with the ensemble. The third row is the MST rank histogram using the CCM3 analyses as verification. They indicate trouble after only one day. The fourth row is when verification is sampled randomly from both NMC and CCM3 analyses. It too indicates trouble after only one day.

In the multi-model context one has a collection of analyses from which verification can be selected. The experiments reported above show that the probabilistic assessment results will be dependent upon the particular verification utilized. Because the collection of multi-model analyses represents some

kind of probabilistic expression of truth, one is in the situation of having both a probabilistic expression for the forecast and for the verification. A first order approach would be to select a deterministic verification at every verification time by sampling randomly from the distribution of analyses. In this case one samples randomly between the CCM3 analyses and the NMC analyses, and the traditional rank histogram assessment produces almost identical results to the NMC-only verification. However, the MST rank histograms are different. Using only the NMC analyses as verification, the MST rank histograms indicate the ensembles break down after a 2 day forecast lead. Drawing randomly from the NMC and CCM3 analyses produces MST rank histograms that indicate that the PM MM ensembles break down after only a 1 day forecast lead (see fourth row of figure 2). This should not be interpreted as the ensemble being poor, but rather that sampling randomly from the ensemble of analyses is not an appropriate method to account for the uncertainty in the verification. Experiments with a toy model support this interpretation. Finding appropriate ways to take the probabilistic verification information into account when assessing PM MM ensembles probabilistically will be a continuing area of research associated with this project.

A larger database of multi-model ensemble forecasts is being obtained from the ECMWF, and a post-doc has (finally) been hired to work on the project.

RESULTS

- It is found that the CCM3 model is a better model of the NMC model than it is of the real weather. This suggests that a successful approach would be to take the analyses from multi-model ensembles, and propagate them forward using a single model, consistent with the results of Richardson (2001).

- It is found that probabilistic assessment of single model ensemble forecasts are dependent upon the verification used. There are quantitative differences between NMC ensemble forecasts that are assessed using NMC analyses, and NMC ensemble forecasts that are assessed using the NMC analyses projected into the CCM3 space. Any analysis is, at best, a projection of truth into the model space. The quantitative impact of projecting one model state into another shown above hints at the impact of projecting truth into different model states.

- Probabilistic assessment of PM MM ensembles gives different answers depending on the choice of verification. Using the true system state, of course, provides the correct assessment, but the true system state is unavailable in NWP. The sensitivity of the projection operation shown for single model ensembles indicates that one should expect similar sensitivity in assessment of the PM MM ensemble, and one does. Treating verification as a random variable further degrades the probabilistic forecast.

IMPACT/APPLICATIONS

If successful, the results of this project will alter the way operational multi-model ensemble forecasts are generated and assessed. Ultimately, it could provide a basis for not only improving existing models, but for intelligently constructing new models whose features optimally supplement existing multi-model ensembles.

TRANSITIONS

None.

RELATED PROJECTS

I am associated with an NSF-funded project that aims to address the impact of model inadequacy in data assimilation and forecasting using a single model structure. Model inadequacy insights gained during the NSF project will be applied to the current project.

REFERENCES

Bishop, C. H., Etherton, B. J., and Mujumdar, S. J. (2001). Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects. Monthly Weather Review, 129:420-436.

Hansen, J. A. (2002). Accounting for model error in ensemble-based state estimation and forecasting. Monthly Weather Review, 130:2373-2391.

Richardson, D. S. (2001). Ensembles using multiple models and analyses. Quarterly Journal of the Royal Meteorological Society, 172:1847-1864.

PUBLICATIONS

None.

PATENTS

None.